# Evidence of Discrimination?

The Department of Developmental Services (DDS) in California is responsible for allocating funds to support over 250,000 developmentally-disabled residents. The data set `ca_dds_expenditures.csv` contains data about 1,000 of these residents. The data comes from a discrimination lawsuit which alleged that California's Department of Developmental Services (DDS) privileged white (non-Hispanic) residents over Hispanic residents in allocating funds. We will focus on comparing the allocation of funds (i.e., expenditures) for these two ethnicities only, although there are other ethnicities in this data set.

There are 6 variables in this data set:

- Id: 5-digit, unique identification code for each consumer (similar to a social security number and used for identification purposes)
- Age Cohort: Binned age variable represented as six age cohorts (0-5, 6-12, 13-17, 18-21, 22-50, and 51+)
- Age: Unbinned age variable
- Gender: Male or Female
- Expenditures: Dollar amount of annual expenditures spent on each consumer
- Ethnicity: Eight ethnic groups (American Indian, Asian, Black, Hispanic, Multi-race, Native Hawaiian, Other, and White non-Hispanic)
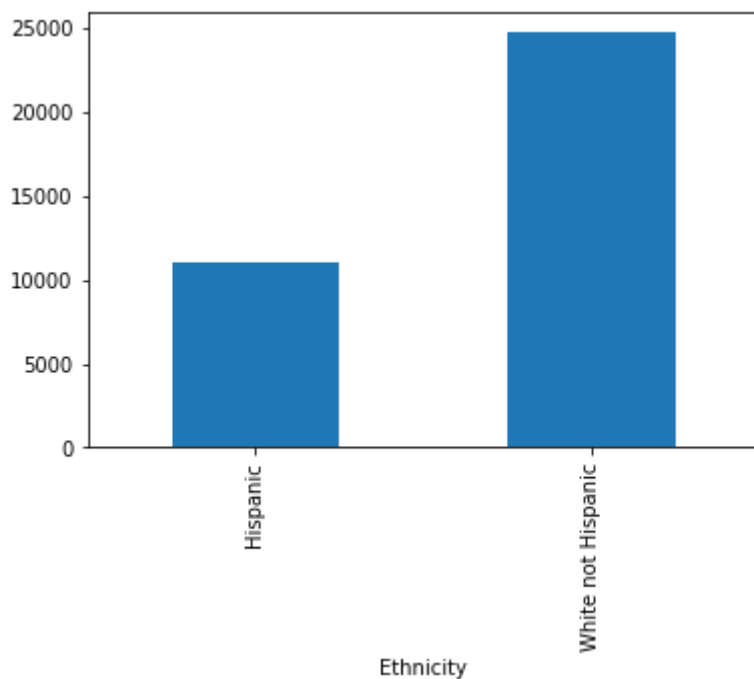
# Question 1

Read in the data set. Make a graphic that compares the *average* expenditures by the DDS on Hispanic residents and white (non-Hispanic) residents. Comment on what you see.

```
In [5]:   # YOUR CODE HERE
          import pandas as pd
          %matplotlib inline
          import matplotlib.pyplot as plt
          import numpy as np

          df=pd.read_csv("ca_dds_expenditures.csv")
          df.head()

          df2=df.groupby(["Ethnicity"]).Expenditures.mean().loc[["Hispanic","White
          not Hispanic"]]
          df2.plot.bar()
```

Out[5]:   <matplotlib.axes._subplots.AxesSubplot at 0x7f99701ebfd0>



**YOUR EXPLANATION HERE**

The average expenditures on "White Non hispanic" are more than "Hispnaic"
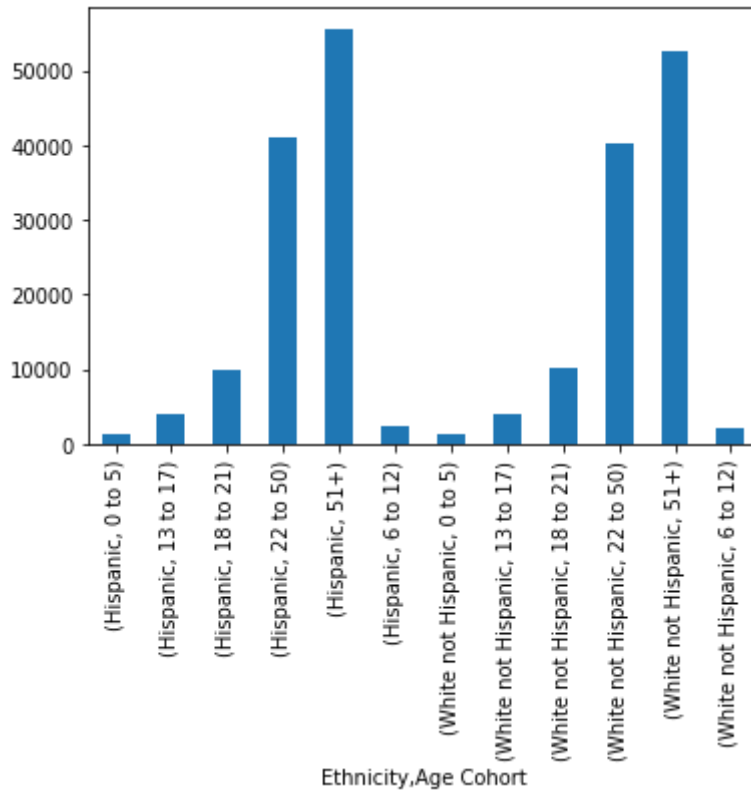
# Question 2

Now, calculate the average expenditures by ethnicity and age cohort. Make a graphic that compares the average expenditure on Hispanic residents and white (non-Hispanic) residents, *within each age cohort*.

Comment on what you see. How do these results appear to contradict the results you obtained in Question 1?

```
In [10]:  # YOUR CODE HERE
          df3 = df.groupby(["Ethnicity","Age Cohort"]
                  ).Expenditures.mean().loc[["Hispanic","White not Hispanic"]]

          df3.plot.bar()
```

Out[10]:  <matplotlib.axes._subplots.AxesSubplot at 0x7f996fa9f518>



**YOUR EXPLANATION HERE**

Average age on each are both ethnicities are the same

# Question 3

Can you explain the discrepancy between the two analyses you conducted above (i.e., Questions 1 and 2)? Try to tell a complete story that interweaves tables, graphics, and explanation.
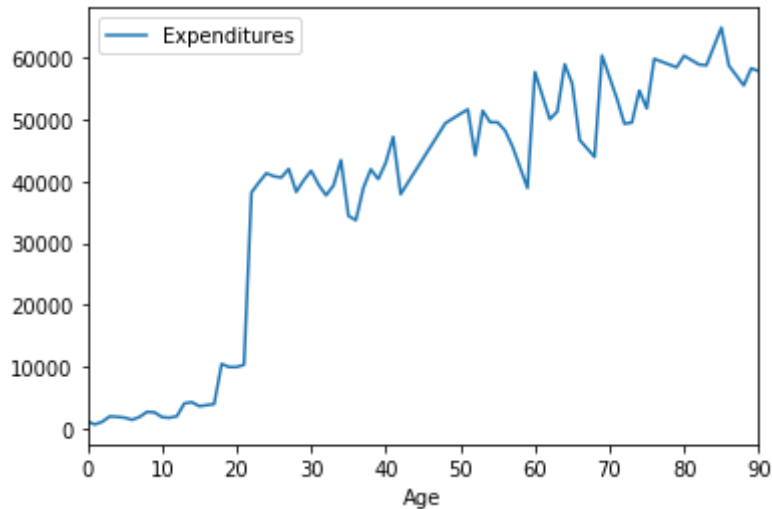
*Hint:* You might want to consider looking at:

- the distributions of ages of Hispanics and whites
- the average expenditure as a function of age

In [11]:
```python
# YOUR CODE HERE (although you may want to add more code cells)
df4=df.groupby(["Ethnicity","Age"]).Expenditures.mean().loc["White not H
ispanic"]

df4.plot(kind='line',alpha=1,legend=True)
```
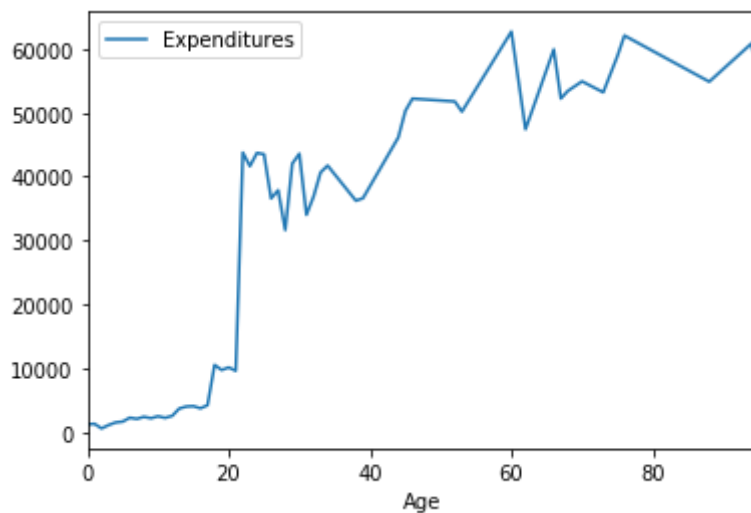
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7f996fa2bcc0>



In [12]:
```python
df4=df.groupby(["Ethnicity","Age"]).Expenditures.mean().loc["Hispanic"]

df4.plot(kind='line',alpha=1,legend=True)
```
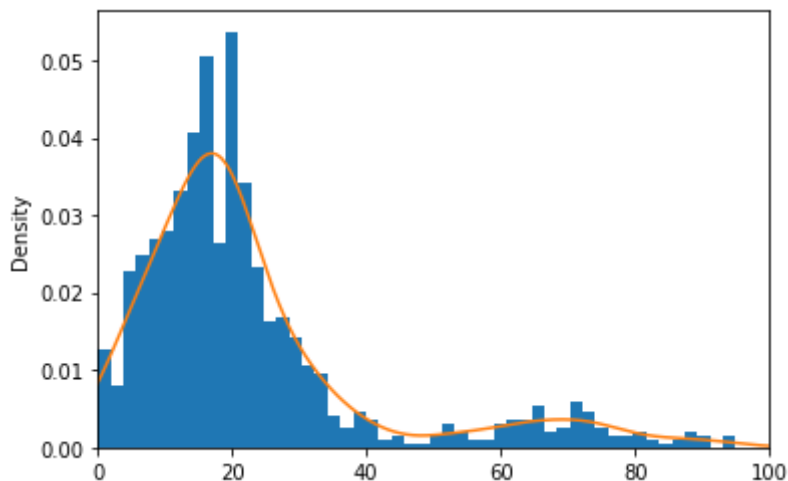
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x7f996f9adda0>

```
In [13]: df.Age.plot.hist(bins=50,density=True)
         df.Age.plot.density(xlim=(0,100))
```

Out[13]: `<matplotlib.axes._subplots.AxesSubplot at 0x7f996f906390>`



**YOUR EXPLANATION HERE (although you may want to add more markdown cells)**

# Submission Instructions

Once you are finished, follow these steps:

1. Restart the kernel and re-run this notebook from beginning to end by going to `Kernel > Restart Kernel and Run All Cells`.
2. If this process stops halfway through, that means there was an error. Correct the error and repeat Step 1 until the notebook runs from beginning to end.
3. Double check that there is a number next to each code cell and that these numbers are in order.

Then, submit your lab as follows:

1. Go to `File > Export Notebook As > PDF`.
2. Double check that the entire notebook, from beginning to end, is in this PDF file. (If the notebook is cut off, try first exporting the notebook to HTML and printing to PDF.)
3. Upload the PDF and Notebook (ipynb) iLearn.